

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein nucleic search, using frame-plus_p2n model

Run on: January 16, 2003, 16:55:57 ; Search time 968.471 Seconds
(without alignments)
330.553 Million cell updates/sec

Title: US-09-856-070-23

Perfect score: 55

Sequence: 1 ELMRLQDYDE 11

Scoring table:
BLASTM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
delop 6.0, delext 7.0

Searched: 2054640 seqs, 14561403878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+2n.m-3; -FV=zip
-Q=/con2_1/USPTO-spool/us09856070/unal_14017003_155844_1621/app_query_tas_a_1_1592
-DB=GenEmbl -QMT=fastcap -SUFFIX=rgt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=3
-UNITS=bits -STAMP=1 -FNU=1 -MATRIX=blosum62 -TRANS=trans40.cdi -LISP=45
-DOCALLIGN=200 -THR_SCORE=100 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=pln -NPM=ext -HEAPSIZ=500 -MINFN=0 -MAXLEN=200000000
-USER=ns09856070.unl_1_155844_1621 -MAIL=14017003_155844_1621 -NPM=6 -FV=zip
-NO_XLPXY -NO_WMAP -LAPQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIME=0-120
-WARN_TIME=0-130 -THRAUD=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -Fdelop=6 -DELEXT=7

Database :

GenEmbl:

- 1: gb_ba:*
- 2: gb_ba:*
- 3: gb_ba:*
- 4: gb_ba:*
- 5: gb_ba:*
- 6: gb_ba:*
- 7: gb_ba:*
- 8: gb_ba:*
- 9: gb_ba:*
- 10: gb_ba:*
- 11: gb_ba:*
- 12: gb_ba:*
- 13: gb_ba:*
- 14: gb_ba:*
- 15: gb_ba:*
- 16: gb_ba:*
- 17: gb_ba:*
- 18: gb_ba:*
- 19: gb_ba:*
- 20: gb_ba:*
- 21: gb_ba:*
- 22: gb_ba:*
- 23: gb_ba:*
- 24: gb_ba:*
- 25: gb_ba:*
- 26: gb_ba:*
- 27: gb_ba:*
- 28: gb_ba:*

- 29: em_vit:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_ror:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	100.0	2514	4	HEVIRINA
2	55	100.0	2930	9	HEMVI13
3	55	100.0	3044	6	AX411074
4	55	100.0	3044	9	HSE2PIN
5	55	100.0	3047	6	AX440476
6	55	100.0	3072	9	HSM802625
7	55	100.0	3173	9	HC013903
8	55	100.0	215705	9	AL589931
9	52	94.5	478	9	AF189213
10	52	94.5	2701	10	MMEZR
11	52	94.5	183362	3	AC087635
12	52	94.5	207782	2	AC074334
13	52	94.5	217917	2	AC022229
14	52	94.5	286508	2	AC156113
15	50	90.9	475	9	AF188897
16	48	87.3	1445	10	AF450294
17	48	87.3	18233	2	AC113744
18	48	87.3	189873	2	AC125736
19	45	81.8	426	9	AF199015
20	42	76.4	182756	2	AF004891
21	42	76.4	187419	2	AF005053
22	42	76.4	233738	9	AC007956
23	41	74.5	136822	5	AC091091
24	40	72.7	243	8	YSALK2B
25	40	72.7	2545	5	AHG19790
26	40	72.7	105596	2	AC099130
27	40	72.7	150928	2	AC103550
28	40	72.7	158179	2	U8JN03136
29	39	70.9	1678	9	HS085994
30	39	70.9	1696	9	HS085995
31	39	70.9	1982	9	HS087408
32	39	70.9	2075	9	AK094451
33	39	70.9	2452	9	AF091361
34	39	70.9	2784	9	AF095771
35	39	70.9	2790	1	SACNTPASE
36	39	70.9	4875	1	SAC437617
37	39	70.9	6583	9	AF082523
38	39	70.9	10079	7	AF155037
39	39	70.9	34831	3	CEC4762
40	39	70.9	50459	4	AF002451
41	39	70.9	95047	2	AC097135
42	39	70.9	95560	2	AC127620
43	39	70.9	96879	9	HS384121
44	39	70.9	96879	9	HS384121
45	39	70.9	96879	9	HS384121

ALIGNMENTS

RESULT 1

HOVEZRINA
 DEFINITION Bos taurus ezrin mRNA, complete cds.
 ACCESSION M98498
 VERSION M98498.1 GI:289407
 KEYWORDS ezrin.
 SOURCE Bos taurus brain cDNA to mRNA
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 2514)
 Bergson,C.M., Zhao,H., Sajioh,K., Human,R.S. and Nestler,E.J.
 Ezrin and osteonectin, two proteins associated with cell shape and
 growth, are enriched in the locus coeruleus
 Mol. Cell. Neurosci. 4, 64-73 (1993)
 JOURNAL
 FEATURES
 source
 1..2514
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /issue_type="brain"
 152..1897
 /codon_start=1
 /product="ezrin"
 /protein_id="AAA30510.1"
 /db_xref="GI:289408"
 /translation="MPKPIINVRITMDAEFAIQNTTGTGKQEDQVVKTIIGLREVWY
 FGLQVYDNGKFPPTWIKLQKVSQAQVKKENPQKFKPAKFPEDVAEELIQDITQKLF
 FLQVKEGILSDIEIGYPPETAVLLGSYAVCAKFGDYNKELIKAGYLSSRELIPQVMDQ
 HKITPQWDRLOVWHABHGMKDSAMIEYIKIAQDLKEMGYNFPEIKKKKGIDML
 GVALGILNYEKDKLTKPIGTWSEIRNISPNDKKKFKVQKIDKAPDFVFPYAPRI
 NKRIQLQCMNHPIYMPKPPPTIPVQCMKAPKPKHQKQLEPQQLTEKKPPPTVF
 PKEQMREKEFLMIPLADYEETKFAEKELSDIQLKLEEFKKAQEAEGLEAD
 RLAAKRAKFEIRQAQDITKQQLATAEYTAIALLKLEAPKPKENVEEMLPRAK
 EAQDLDLTKRELLHVMTPAPPVPEVYVNVHIEGPGUEGTELSAEISGLSDSN
 EFKRIEAKNFKVQRLMTLISELSQARDENKKTNDIHNENMROGRKRYKTKIROL
 KQNTKQRIIDEFAE"
 BASE COUNT 644 a 658 c 738 g 474 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0154 Length: 2514
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0
 US-09-856-070-23 (1-11) x HOVEZRINA (1-2514)
 QY 1 GluteuMetLeuArqLeuGlnAspTyrGluGlu 11
 Db 1187 GAGTGAIGTCAGGGTTTCAGSACIAAGSAA 1219
 RESULT 2
 HUMVIL2
 LOCUS Human cytovillin 2 (VIL2) mRNA, complete cds.
 DEFINITION J05021
 ACCESSION J05021
 VERSION J05021.1 GI:340216
 KEYWORDS cytovillin; cytovillin 2; microvillar protein.
 SOURCE Human placenta, cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 2430)
 Turunen,O., Winqvist,R., Pakkanen,R., Grzeschik,K.H., Wahlstrom,T.
 and Vaheri,A.
 Cytovillin, a microvillar Mr 75,000 protein, cDNA sequence,
 prokaryotic expression, and chromosomal localization
 J. Biol. Chem. 264 (28), 16727-16732 (1989)
 JOURNAL 89380299
 MEDLINE
 PubMed 2674140

COMMENT Draft entry and computer-readable sequence for 11 kindly submitted
 by O.Turunen, 31-AUG-1989.
 FEATURES
 source
 1..2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="6q22 q27"
 1..2930
 /gene="VIL2"
 110..1837
 /gene="VIL2"
 /note="cytovillin 2"
 /codon_start=1
 /protein_id="AAA61278.1"
 /db_xref="GI:340217"
 /db_xref="GDB:600-120-489"
 /translation="MDAIEFAIQNTTGTGKQEDQVVKTIIGLREVWYFGLIYVNDKGF
 PIWKIDKKVSQAQVKKENPQKFKPAKFPEDVAEELIQDITQKLF
 FLQVKEGILSDIEIGYPPETAVLLGSYAVCAKFGDYNKELIKAGYLSSRELIPQVMDQ
 HKITPQWDRLOVWHABHGMKDSAMIEYIKIAQDLKEMGYNFPEIKKKKGIDML
 GVALGILNYEKDKLTKPIGTWSEIRNISPNDKKKFKVQKIDKAPDFVFPYAPRI
 NKRIQLQCMNHPIYMPKPPPTIPVQCMKAPKPKHQKQLEPQQLTEKKPPPTVF
 PKEQMREKEFLMIPLADYEETKFAEKELSDIQLKLEEFKKAQEAEGLEAD
 RLAAKRAKFEIRQAQDITKQQLATAEYTAIALLKLEAPKPKENVEEMLPRAK
 EAQDLDLTKRELLHVMTPAPPVPEVYVNVHIEGPGUEGTELSAEISGLSDSN
 EFKRIEAKNFKVQRLMTLISELSQARDENKKTNDIHNENMROGRKRYKTKIROL
 KQNTKQRIIDEFAE"
 BASE COUNT 793 a 658 c 821 g 658 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0183 Length: 2930
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-856-070-23 (1-11) x HUMVIL2 (1-2930)
 QY 1 GluteuMetLeuArqLeuGlnAspTyrGluGlu 11
 Db 1112 GAGTGAIGTCAGGGTTTCAGSACIAAGSAA 1144
 RESULT 3
 AX411074
 LOCUS AX411074
 DEFINITION Sequence 3721 from Patent WO229103.
 ACCESSION AX411074
 VERSION AX411074.1 GI:21443779
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1
 Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
 Gene expression profiles in liver cancer
 Patent: WO 229103-A 3721 11-APR 2002;
 GENE LOGIC INC (US)
 FEATURES
 source
 1..3044
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="EMBL/GenBank Accession No. X51521"
 BASE COUNT 826 a 687 c 855 g 675 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0191 Length: 3044
 Score: 55.00 Matches: 11
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6q22-q27"
/clone="DKFZp762H157"
/tissue_type="melanoma (McMv cell line)"
/clone_lib="762 (SYM:ym: hmc12). Vector pSPori1, host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..1894
/clone="DKFZp762H157"
1..1894
/clone="DKFZp762H157"
/notes="ezrin"
/product="hypothetical protein"
/protein_id="CAH82418.1"
/db_xref="GI:7428175"
/translation="MPKINVRVTMTAELEFAIQNTTIGKQVVKITIGLREVWYFGLHYVNDKGFPTW
LKDRKVSADVFPPNPLPFFKFAKFFVFVAEELIADITKLLFLAVKEGILSDIY
CPDFAVILGSAVQAKFGVWKFVKKSTY SSPI 190PMVGHKI TPDQWPIQV
WHAHRCMIKDAMLELYIKIADLEMYGIVNFEIKNKKGTIDIWGVNAGIINYEKID
KIDPKIGFWSELRNISENKKFVIKPIUKKAPDVFVAPRLRNKRLIOLCMONHEL
YMRKRPDIEVQKKAARBEKHQKLERQLETEKKRRETVREKEMRREELM
LRQDTEKKKAERELSPQIQALQLEFEKRPAGFPAPLEADPMALPAKRELEPQ
AVDTKSPQIAAFTAYTAKTATILEFAFPKDEVEEMHEAKAQQDLVYKKEELH
LVMTAPPPPPVYFVSFVQVSEIQDAGAPFVGYSAFLSSEGIKIDRNEKKRI TEAK
KNRQVQQLITLSSELISQARDENKKTNDIHNENMRQCKRYKTLIQIKQCNIKQRI
DEPAL"
3032..3037
polya_signal
polya_site
BASE COUNT 846 a 688 c 868 g 670 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0193 Length: 3072
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-856-070-23 (1-11) x USM802625 (1-3072)
QY 1 GlutLeuMetLeuArgLeuGlnAspTyrGluGlu 11
|||||
Db 1169 GAGTTGATGCGGGGTCGAGGACTATGAGGAG 1201
RESULT 7
LOCUS BC013903 3173 bp mRNA linear PRI 10-SEP-2001
DEFINITION Homo sapiens, Similar to villin 2 (ezrin), clone MCC:1584
IMAGE:2959399, mRNA, complete cds.
ACCESSION BC013903.1 GI:15530242
VERSION MGC.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;
Strausberg,K.
Submitted (07-SEP-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A04, Bethesda, MD 20892-2540,
USA
Contact: MGC help desk
Email: gcapbs-r@mail.nih.gov
Tissue Procurement: AICC
cDNA Library Preparation: Pabin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

```

DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbio.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/Link at: http://image.llnl.gov
Series: IRAL Plate: 2 Row: a Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9257254.
FEATURES
Location/Qualifiers
source
1..3173
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MCC:1584 IMAGE:2959399"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B-R"
/vector="Vector, pOTB7"
142..1902
/codon_start=1
/product="Similar to villin 2 (ezrin)"
/protein_id="AAH13903.1"
/db_xref="GI:15530243"
/translation="MPKINVRVTMTAELEFAIQNTTIGKQVVKITIGLREVWY
FGLHYVNDKGFPTW LKDRKVSADVFPPNPLPFFKFAKFFVFVAEELIADITKLLFLAVKEGILSDIY
CPDFAVILGSAVQAKFGVWKFVKKSTY SSPI 190PMVGHKI TPDQWPIQV
WHAHRCMIKDAMLELYIKIADLEMYGIVNFEIKNKKGTIDIWGVNAGIINYEKID
KIDPKIGFWSELRNISENKKFVIKPIUKKAPDVFVAPRLRNKRLIOLCMONHEL
YMRKRPDIEVQKKAARBEKHQKLERQLETEKKRRETVREKEMRREELM
LRQDTEKKKAERELSPQIQALQLEFEKRPAGFPAPLEADPMALPAKRELEPQ
AVDTKSPQIAAFTAYTAKTATILEFAFPKDEVEEMHEAKAQQDLVYKKEELH
LVMTAPPPPPVYFVSFVQVSEIQDAGAPFVGYSAFLSSEGIKIDRNEKKRI TEAK
KNRQVQQLITLSSELISQARDENKKTNDIHNENMRQCKRYKTLIQIKQCNIKQRI
DEPAL"
930 a 691 c 868 g 684 t
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.: 0.02 Length: 3173
Score: 55.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-856-070-23 (1-11) x BC013903 (1-3173)
QY 1 GlutLeuMetLeuArgLeuGlnAspTyrGluGlu 11
|||||
Db 1177 GAGTTGATGCGGGGTCGAGGACTATGAGGAG 1209
RESULT 8
LOCUS AL589931 215705 bp DNA linear PRI 15-SEP-2001
DEFINITION Human DNA sequence from clone pB11-507010 on chromosome 6q25.2-26,
complete sequence.
ACCESSION AL589931 ACC24785
VERSION AL589931.14 GI:15591370
KEYWORDS HIG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;
Sehrai,H.
Submitted (15-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonesrequest@sanger.ac.uk
on Sep 12, 2001 this sequence version replaced gi:14586248.
During sequence assembly data is compared from overlapping clones.

```



```

* 187911 196625: contig of 2715 bp in length
* 196626 196645: gap of unknown length
* 192088 192088: contig of 1443 bp in length
* 192089 192198: gap of unknown length
* 192104 193647: contig of 1529 bp in length
* 193648 193657: gap of unknown length
* 195040 195040: contig of 1383 bp in length
* 195041 195060: gap of unknown length
* 195061 196945: contig of 1885 bp in length
* 196946 196965: gap of unknown length
* 196966 198219: contig of 1254 bp in length
* 198220 198249: gap of unknown length
* 198249 199607: contig of 1668 bp in length
* 199608 199927: gap of unknown length
* 201164 201164: contig of 1236 bp in length
* 201164 202937: gap of unknown length
* 202937 202937: contig of 844 bp in length
* 202938 202947: gap of unknown length
* 202948 202952: contig of 855 bp in length
* 202953 202953: gap of unknown length
* 202953 204150: contig of 1228 bp in length
* 204151 204170: gap of unknown length
* 204171 205707: contig of 1537 bp in length
* 205708 205727: gap of unknown length
* 205728 207782: contig of 2055 bp in length.
FEATURES
    source
        1..207782
            /organism="Mus musculus"
            /strain="C57BL6/J"
            /db_xref="taxon:10090"
            /chromosome="15"
            /clone="RP23-103120"
            /sex="male"

```

```

BASE COUNT  54791 a 49882 c 48733 g 53275 t 1101 others
ORIGIN

```

```

Alignment Scores:
Pred. No.:      10.6      Length:      207782
Score:          52.00     Matches:      10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 90.91% Mismatches: 0
Query Match:      94.55% Indels:      0
db:                2      Gaps:        0

```

```

US-09 856-070-23 (1-11) x AC024329 (1-207782)

```

```

CY 1 CluLeuMetLocArgLeuGlnAspIlyrC18C13 11
|||||
db 174047 GAGGTGATGCTGGCTGCAAGATAGAGAG 174015

```

```

RESULT 13
AC024329/c
1000S
DEFINITION
    Mus musculus clone RP23-103120 strain C57BL6/J, WORKING DRAFT
SEQUENCE
    AC024329
AC024329
AC024329.8 C1:18057060
HIG: HIGS_PHASE1: HIGS_DRAFT.
SOURCE
    Mus musculus
    ORGANISM
        Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
        1 (bases 1 to 217917)
        /organism="Mus musculus"
        /strain="C57BL6/J"
        /db_xref="taxon:10090"
        /clone="RP23-103120"
        /sex="male"
REFERENCE
    AUTHORS
        Montgomery, K. J., Shim, C., Decker, J., Thomas, P., Perera, A.,
        Ioshikhes, I. P., Grills, C., Han, J., Lee, P., Long, J.,
        Gordon, M., Golitz, J. S. and Kucherlapati, R.
        High Throughput Mouse Sequencing
        Unpublished
        2 (bases 1 to 217917)
        /organism="Mus musculus"
        /strain="C57BL6/J"
        /db_xref="taxon:10090"
        /clone="RP23-103120"
        /sex="male"
REFERENCE
    AUTHORS
        Montgomery, K. J., Grills, C., Han, J., Lee, P., Long, J.,
        Ioshikhes, I. P., Shim, C., Decker, J., Thomas, P., Perera, A.,
        Gordon, M., Golitz, J. S. and Kucherlapati, R.
        Direct Submission

```

JOURNAL

COMMENT:

```

Submitted (01 Feb 2003) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Jun 4, 2002 this sequence version replaced gi:11044413.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpgc.org/sequence/mouse.html
Contact: hpgc@mcml.mgh.harvard.edu
-----Summary Statistics
Center project name: AAC
Sequencing vector: pBC18; L08752
Chemistry: Dye terminator Big dye, 100%
*Consensus quality: 212564 at least Q20
*Consensus quality: 211720 at least Q40
*Consensus quality: 210397 at least Q40
**Estimated insert size: agarose FF: N/A
**Estimated insert size: 217717 - sum-of-continqs
Quality coverage: agarose FF: N/A
Quality coverage: 8.5 x in Q20 bases; sum-of-continqs estimation
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 60900: contig of 60900 bp in length
* 60901 60920: gap of unknown length
* 106520 106530: contig of 45600 bp in length
* 106540 106549: gap of unknown length
* 106541 142322: contig of 35782 bp in length
* 142323 142342: gap of unknown length
* 142343 176177: contig of 33835 bp in length
* 176178 176197: gap of unknown length
* 176198 204272: contig of 28075 bp in length
* 204273 204292: gap of unknown length
* 204293 209252: contig of 4960 bp in length
* 209253 209272: gap of unknown length
* 209273 214332: contig of 5060 bp in length
* 214333 214352: gap of unknown length
* 214353 215731: contig of 1379 bp in length
* 215732 215751: gap of unknown length
* 215752 215800: contig of 49 bp in length
* 215801 215820: gap of unknown length
* 215821 216675: contig of 855 bp in length
* 216676 216695: gap of unknown length
* 216696 217917: contig of 1222 bp in length.

```

FEATURES

```

    source
        1..217917
            /organism="Mus musculus"
            /strain="C57BL6/J"
            /db_xref="taxon:10090"
            /clone="RP23-103120"
            /sex="male"
        1..60900
            /note="assembly_name:Contig211"
            /clone_end="896"
            /note="assembly_name:Contig210"
            /vector_side="left"
            /note="assembly_name:Contig209"
            /note="assembly_name:Contig208"
            /clone_end="17"
            /vector_side="left"
            /note="assembly_name:Contig207"
            /note="assembly_name:Contig206"

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature 209273..214332
 /note="assembly_name:Contig205"
 misc_feature 214333..215731
 /note="assembly_name:Contig204"
 misc_feature 215732..215900
 /note="assembly_name:Contig203"
 misc_feature 215901..215975
 /note="assembly_name:Contig202"
 misc_feature 215976..217917
 /note="assembly_name:Contig201"
 BASE COUNT 59588 a 52987 c 4774 g 5531 t 227 others
 ORIGIN

Alignment Scores:
 Pred. No.: 11.1 length: 217917
 Score: 52.00 Matches: 10
 Percent Similarity: 100.00% Conservatve: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 94.55% Indels: 0
 DB: 2 Gaps: 0

US-09-856-070-23 (1-11) x A02223294 (1-217917)

QY 1 GlutamineLeuArgLeuGlnAspTyrGluGlu 11

DB 10729 GAGTATATCTTGGATGAGAGATACAGATG 10697

RESULT 14
 AC125143 286508 bp DNA Linear HTG: OR AUG 2002
 LOCUS Mus musculus chromosome YNK clone pp24-342H13, WAPPING DRAFT
 DEFINITION SEQUENCE, 10 unordered pieces.
 ACCESSION AC125143
 VERSION AC125143.1 GI:22138589
 KEYWORDS HTG; HTGS; PHASE1; HTCS; DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi,
 Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Mus.
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 286508)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (08-AUG-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@watson.wustl.edu
 Project Information
 Center project name: M.BB0342H13
 Summary Statistics

Sequencing vector: M13; 0%
 Sequencing method: plasmid; 100%
 Chemistry: Dye-terminator ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: phrap, version 0.99011a
 Consensus quality: 283924 bases at least Q40
 Consensus quality: 285778 bases at least Q30
 Consensus quality: 286826 bases at least Q20
 Insert size: 225000, average 1p
 Insert size: 304647, sum-of-coverage
 Quality coverage: 14.29 in Q20 bases; average 1p
 Quality coverage: 8.54 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
 consists of 10 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 be preserved.
 1 1503: contig of 1503 bp in length
 1504: gap of unknown length
 1604 4585: contig of 2982 bp in length
 4586: gap of unknown length
 4587: contig of 18201 bp in length
 4588: gap of unknown length
 4589: contig of 21961 bp in length
 4590: gap of unknown length
 4591: contig of 31906 bp in length
 4592: gap of unknown length
 4593: contig of 26715 bp in length
 4594: gap of unknown length
 4595: contig of 56889 bp in length
 4596: gap of unknown length
 4597: contig of 123560 bp in length
 4598: gap of unknown length
 4599: contig of 851 bp in length
 4600: gap of unknown length
 4601: contig of 1040 bp in length.
 Location/Qualifiers
 1..286508
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chr_m_sname="YNK"
 /chr_m_slocus="pp24-342H13"
 1..11503
 /note="assembly_name:Contig99"
 1604..4585
 /note="assembly_name:Contig100"
 4586..4586
 /note="assembly_name:Contig101"
 22387..44947
 /note="assembly_name:Contig102"
 45048..76953
 /note="assembly_name:Contig103"
 77054..103768
 /note="assembly_name:Contig104"
 103869..160757
 /note="assembly_name:Contig105"
 160858..284417
 /note="assembly_name:Contig106"
 284518..285368
 /note="assembly_name:Contig107"
 285469..286508
 /note="assembly_name:Contig108"
 286509..286508
 /note="assembly_name:Contig109"
 286509..286508
 /note="assembly_name:Contig110"

FEATURES

SOURCE

BASE COUNT 78361 a 64449 c 65769 g 77032 t 907 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 15.1 length: 286508
 Score: 52.00 Matches: 10
 Percent Similarity: 100.00% Conservatve: 1
 Best Local Similarity: 90.91% Mismatches: 0
 Query Match: 94.55% Indels: 0
 DB: 2 Gaps: 0
 US-09-856-070-23 (1-11) x A02223294 (1-286508)
 QY 1 GlutamineLeuArgLeuGlnAspTyrGluGlu 11
 DB 129689 GAGTATATCTTGGATGAGAGATACAGATG 10697
 RESULT 15
 AC188897
 AC188897 AF188897 475 bp mRNA linear PRI 19-OCT-1999
 LOCUS Homo sapiens ovary early mRNA, partial cds.
 DEFINITION
 ACCESSION

```

VERSION      AF188897.1  GI:6063144
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
               Homo sapiens.
REFERENCE     1. (bases 1 to 475)
AUTHORS      Chen, Z.C., Fadiel, A. and Naftolin, F.
TITLE        Ezrin gene mutation in ovarian cancer
JOURNAL      Unpublished
REFERENCE     2. (bases 1 to 475)
AUTHORS      Chen, Z.C., Fadiel, A. and Naftolin, F.
TITLE        Direct Submission
JOURNAL      Submitted (22 SEP 1999) DB/GYN, Yale University, 333 Cedar Street,
               New Haven, CT 06511, USA

FEATURES
  source
    1..475
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="6"
      /map="6q22-q27"
      /cell_line="Bix1"
      /cell_type="epithelial cancer cells"
      /tissue_type="ovary"
    <1..475
      /note="villin 2; Kinase substrate"
      /codon_start=2
      /product="ezrin"
      /protein_id="MAF03155.1"
      /db_xref="GI:6063145"
      /translation="TGMVGINFVKKKKGGIDWGVIALGLNIYKDDKLXPKIGPP
      WSEIKNISFNKKKPVINDKNAFDFVYAFGLKINKKIDQICMGNHLYMKKKKPD
      IEVQMKAAAREEKQLEPQQLTEKKPEXVERPKFQMPREKEELMLPLQDYE"
BASE COUNT   154 a  90 c  136 q  93 t  2 others
ORIGIN

Alignment Scores:
  Prod. No.:      0.0341      Length:      475
  Score:          50.00      Matches:      10
  Percent Similarity: 100.00%      Conservative: 0
  Best Local Similarity: 100.00%      Mismatches: 0
  Query Match:     90.91%      Indels:      0
  DB:              9          Gaps:          0

US 09-856-070-23 (1 11) x AF188897 (1-475)
Qy  1 GluLeuMetLeuArgLeuGluAspTyrGlu 10
    | | | | | | | | | | | | | | | | | | | |
Db  446 GACITGATCTTGGGGGCTGACGACATCAAA 475

Search completed: January 16, 2003, 19:05:54
Job time : 1034.47 secs

```